# A Survey of the Coverage and Methodologies of Schemas and Vocabularies Used to Describe Information Resources

Philip Hider

School of Information Studies, Charles Sturt University, Boorooma Street, Wagga Wagga, NSW 2678, Australia, <phider@csu.edu.au>

Philip Hider has been Head of the School of Information Studies at Charles Sturt University in Australia since 2008. He holds a Master of Librarianship degree from the University of Wales, Aberystwyth, and a PhD from City University, London. He worked at the British Library from 1995-1997 and in Singapore from 1997-2003. He specialises in the area of knowledge organization, and served on the Australian Committee on Cataloguing from 2004-2009. He is a fellow of the Chartered Institute of Library and Information Professionals and an associate member of the Australian Library and Information Association.

Hider, Philip. **A Survey of the Coverage and Methodologies of Schemas and Vocabularies Used to Describe Information Resources.** *Knowledge Organization. 42(3)*, 154-163. 24 references.

**Abstract:** Riley's survey (2010) of metadata standards for cultural heritage collections represents a rare attempt to classify such standards, in this case according to their domain, community, function and purpose. This paper reports on a survey of metadata standards with particular functions, i.e. those of schemas and vocabularies, but that have been published online for any domain or community (and not just those of the cultural heritage sector). In total, 53 schemas and 328 vocabularies were identified as within scope, and were classified according to their subject coverage and the type of warrant used in their reported development, i.e. resource, expert or user warrant, or a combination of these types. There was found to be a general correlation between the coverage of the schemas and vocabularies. Areas of underrepresentation would appear to be the humanities and the fine arts, and, in the case of schemas, also law, engineering, manufacturing and sport. Schemas would appear to be constructed more by consulting experts and considering end-users' search behaviour; vocabularies, on the other hand, are developed more by considering the information resources themselves, or by combining a range of methods.

Received: 22 February 2015; Revised: 4 June 2015; Accepted: 29 June 2015

Keywords: warrant, vocabularies, schemas, standards, metadata, information resources

## 1.0 Introduction

This paper was inspired by Riley's colourful chart, shown in Figure 1, of the "metadata universe" as it existed in 2010. The chart provides a visualisation of relationships between the many metadata standards applied in the cultural heritage sector (so perhaps more of a metadata galaxy than the entire universe). As the metadata universe continues to evolve, it is worth taking additional snapshots. This paper reports on another snapshot, taken through a slightly different lens. Whereas Riley's survey (2010) focused on standards relating to cultural heritage resources, this snapshot covers information resources more generally. Conversely, while Riley covered standards for a wide range of purposes (preservation metadata, rights metadata, structural metadata, etc.), we are primarily interested here in those that pertain to what Riley calls "descriptive metadata," or what sometimes is called "discovery metadata." That is, the survey reported in this paper focuses on metadata standards intended to support information access, or, to be more precise, information resource access, as, like Riley, we are not concerned here with metadata in the more general, structured data sense.

Riley's snapshot (2010) provides an overview of the domains, communities and functions, as well as purposes, which the various cultural heritage standards support. Again, we focus here specifically on two of the functions identified by Riley, i.e. those of structure standards (used

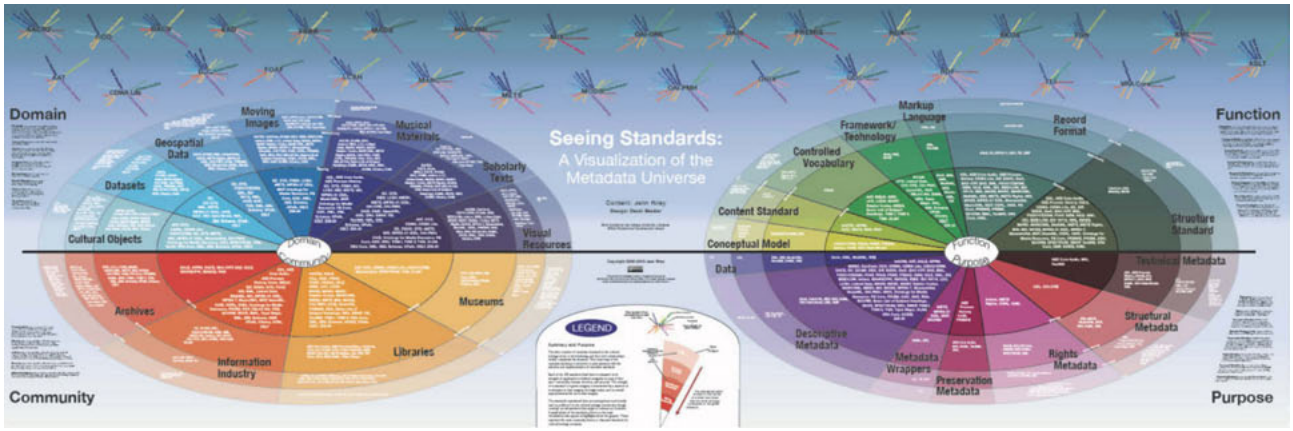


*Figure 1.* A visualization of the metadata universe (Riley 2010; http://www.dlib.indiana.edu/~jenlrile/metadatamap)

to describe information resources as a whole) and of controlled vocabularies (used to describe particular aspects of information resources). An example of the former would be the cataloguing code, *Resource Description and Access,* and of the latter, *Library of Congress Subject Headings*. In this paper, these standards shall be referred to as schemas and vocabularies respectively.

Although Riley's chart offers a useful classification, it is perhaps aimed more at the practitioner, who needs to choose a suitable standard or set of standards—for use by a particular community, to serve a particular function, etc. The classification is relative, grouping standards into particular categories, but not showing gaps, i.e. areas that lack standards. To obtain a picture of how metadata standards are distributed across all information domains and communities, they need to be viewed against the backdrop of the information universe, or at least against a proxy for it—the *Dewey Decimal Classification* has been used for this study. The respective coverage of the schemas and vocabularies identified for this paper can then be compared. Given the articulated nature of the two functions, with vocabularies feeding into schemas, one might expect a certain degree of correlation, although other factors may work against this.

As well as looking at the coverage of schemas and vocabularies, across domains and communities, this paper also considers their reported development methodology. Although the principles of the construction of controlled vocabularies have been expounded, and debated, for a very long time, details of how these principles are to be operationalised are rarely provided in standard guides (at least not in the LIS field), and it is unclear which principles, if any, are emphasised in the development of actual vocabularies. In the case of schemas, even the principles are unclear: there are some accounts of how particular schemas have been developed, but little discussion about how they should be developed. Again, the development methodology of schemas and vocabularies will be compared using a broad taxonomy based on the concept of warrant. Whether the methodologies used were the best ones, however, is a question beyond the scope of this paper.

## 2.0 Literature Review

This section reports on the reviews of two related parts of the literature, namely that which focuses on metadata standards in general and that which focuses on their construction and development. An examination of the latter is necessary in order to construct a framework for the second element of the analysis being reported, for which no established framework presently exists.

### *2.1 Metadata standard surveys and aggregations*

Riley's map of metadata standards appears to be, rather surprisingly, unique. Indeed, few compilations, or bibliographies, of metadata standards have attempted any form of analysis of them; instead, professionals are supported by databases and lists, typically operating as registries. Examples of registries, or registers, include: the Basel Register of Thesauri, Ontologies & Classifications (BARTOC; bartoc.org); the Dublin Core Metadata Registry (http://dcmi.kc.tsukuba.ac.jp/dcregistry); and the Open Metadata Registry (http://metadataregistry.org). Established lists and databases of vocabularies and related standards include: the American Society for Indexing List of Online Thesauri and Authority Files (www.asindexing.org/about-indexing/thesauri/online-thesauri-and-authority-files); the Finnish Ontology Library Service (ONKI; https://onki.fi); Schema.org (https://schema.org); the Society of American Archivists Metadata Directory (http://www2.archivists.org/groups/metadata-and-digital-object-roundtable/metadata-directory); Taxonomy Warehouse (taxonomywarehouse.com); and Vocabulary Bank (http://culturegrid.

lexaurus.net). These compilations serve diverse communities and include a wide range of standards, not all of which are knowledge organisation standards in the narrower sense of information resource description.

The literature about these compilations tends to focus on their operation (e.g. Baker et al., 2002; Johnston 2004), although this does include some discussion around their intellectual organisation: for instance, Souza, Tudhope and Almeida (2012) propose a taxonomy of knowledge organisation systems (KOS). Perhaps the most extensive survey of metadata standards, which includes discussion of their respective merits, is to be found in Zeng and Qin (2008). Although such texts break down standards in various ways, typically by function and field, they do not constitute what might be termed an environmental scan. That is, their aim is to describe and explain, rather than to analyse how the universe of metadata standards fits, or does not fit, together. There is no survey specifically concerned with the way standards have been developed.

### *2.2 Development methodology*

The concept of warrant was chosen as the lens through which to compare how the standards collected for this study had been developed. Kwaśnik (2010, 106) explains how a typology of warrant "provides us with a set of conceptual tools that can be used to understand, analyse, evaluate and design any knowledge-representation system." Comparing the design of different systems can thus be carried out with reference to this core concept. Although other lenses could have been used for framing the comparison, warrant afforded for a relatively simple and straightforward typology that could be readily and consistently applied, without being associated with one particular kind of knowledge-representation system.

Beghtol's identifies (1986) four types of semantic warrant for bibliographic classification systems, that is, bases on which library classification schemes could be constructed: literary, philosophical/scientific, educational and cultural. Although Howarth and Jansen (2014) have suggested that other types of warrant may also be applied, the four approaches identified by Beghtol appear prominently in guides to the development of controlled vocabularies, including, but not limited to, classification schemes. For example, Aitchison, Gilchrist and Bawden (2000) recommend collecting concepts, as well as terms, from: reference works and experts" experience and knowledge (i.e. educational and scientific warrant); the literature (i.e. literary warrant); and, search logs and users' experience and knowledge (i.e. a form of cultural warrant). Theoretically, one can see how the typology supports basic functions of controlled vocabularies: they need to a) represent users' queries (cultural or user warrant); b) represent the resources available (literary or resource warrant); and, c) help users, particularly inexpert ones, identify and articulate their information needs (educational or scientific warrant).

The term warrant was first used by Hulme a century ago, when librarians debated whether bibliographic classification schemes should be developed on an *a priori,* philosophical basis, or on an *a posteriori* basis, that is, according to what Hulme called "literary warrant" (Hulme 1911; 1912; Rodriguez 1984). Advocates of schemes based on a predetermined ontology assumed that libraries reflected, at least collectively, an approximation of knowledge at large. Hulme argued that the knowledge to be found" in books would be more accurately represented by examining the books themselves, and constructing classes when they were warranted by the literature. The practical nature of this approach became increasingly apparent as collections grew and the knowledge they conveyed both expanded and changed, and as subject description became somewhat more exhaustive and specific with the introduction of subject headings. Concepts and terms were derived through a systematic examination of the item in hand. Although focused on textual matter originally, literary warrant could be, and eventually was, applied to a wide range of information resources; I use the term resource warrant to represent the broader application.

The *a priori* approach was not altogether abandoned, however. Indeed, it enjoyed something of a revival through the work of Bliss, Ranganathan, the Classification Research Group, and others, in the middle decades of the twentieth century. Sayers (1967) characterises their approach as "neo-classical;" it was underpinned by a view of knowledge that was less monolithic than that of the nineteenth classificationists, but that was still essentially objective. Knowledge was a cumulative, and mostly unilinear, product of science, or, in the case of the arts and humanities, an educational consensus (Bliss 1939). Thus it could be classified by consulting scientific and academic authorities. This appeal to authority is characteristic of the *a priori* approach in general—the first principles must come from somewhere. Various authoritative sources, such as reference works and domain experts, are identified and utilised. The domain-specific classification schemes and thesauri compiled from the mid-twentieth century onwards emphasised this form of warrant (many of the early ones were compiled by those affiliated with the Classification Research Group). For the purposes of this survey, the concepts of philosophical, scientific and educational warrant are collapsed under the superordinate "expert warrant."

Beghtol's fourth type of warrant, i.e. cultural, is harder to attribute to particular theorists; however, part of this warrant would appear to pertain to the end-users of vo-

cabularies, that is, to their own ways of thinking about, and describing, knowledge (Howarth and Jansen, 2014). The importance of the user perspective was increasingly recognised by vocabulary builders through the second half of the twentieth century, due in part to the increasing accommodation users were given to drive automated retrieval systems. It also became easier to capture this perspective, e.g. through transaction logs. Other methods based on this approach include surveying end-users directly and analysing reference queries. Even the Classification Research Group recognised the value of what they termed enquiry warrant (Howarth and Jansen, 2014). Lancaster (1986) contrasts user warrant and literary warrant, and regards the two as complementary. Similarly, the ISO standard, 25964-1, *Thesauri for Information Retrieval* (2011) advises that consideration is given both to the materials to be indexed and to what "the users want to search for." Previously, Kim and Kim (1977) had questioned whether the two approaches do, in fact, yield particularly different results, and did not find significant differences in their study, but theirs was a rare dissenting voice.

Employing a combination of resource, expert and user warrant is typically advised in the guides to the construction of thesauri and other controlled vocabularies, even if the *way* in which they should be combined is rarely clarified (Mai 2008). Often a list of sources is simply suggested. For instance, Broughton (2006) advises that key terms and concepts in a field be identified by consulting reference sources and expert opinion, and by examining samples of published literature and relevant collections. As noted above, Aitchison, Gilchrist and Bawden (2000) recommend a similar list. For the slightly different purpose of web information architecture, Morville and Rosenfeld (2007) advise that "labels" (for website menus, etc.) should be based on content analysis, and by consulting authors, users (e.g. through card-sorting and free-listing exercises, search log and/or tag analysis) and subject experts.

The general implication of these recommendations is that concepts established by multiple approaches should be favoured, but in cases of conflict, little advice is forthcoming. Lancaster (1986, #) suggests that having users select terms from the literature "is probably the best approach of all ... In this way, the requirements for literary and user warrant are satisfied in a single step." This proposition remains untested, however; it is by no means clear that terms and concepts derived from a single source are not worthy of inclusion, nor that expert warrant can be ignored.

Advice is even less forthcoming when it comes to the development of schemas. Warrant does not appear to have been discussed at all in this context. Typically, case studies report the use of a combination of methods, without explicit justification. For example, Riley and Dalmau (2007) describe expert input and user studies in the construction of a schema for a digital sheet music collection. The methods reported are often similar to those employed by vocabulary builders, although comparisons are not made.

## 3.0 Survey Design

Those schemas and vocabularies that are freely accessible online were targeted for the survey. Following Hider (2012) and Miller (2011), a schema is defined here as a data structure used to describe an information resource, whereas a vocabulary is defined as a set of controlled values used to describe a particular element of a schema, e.g. subject. Vocabularies may well indicate semantic relationships between values, but do not have to; examples of vocabularies include thesauri, subject headings, taxonomies, classification schemes, and some ontologies. Standards about or for schemas and vocabularies (e.g. concerning their construction or display) were not included in the sample. The schemas and vocabularies were searched for using relevant registries and directories, including those cited above, as well as standard search engines (e.g. Google). The search ceased when the two research associates, working semi-independently, had used up all their available hours for the project, which totalled 346. At this point, new finds were occurring quite infrequently—the sample is thus considered to be representative (at the time of the search), though not exhaustive. Those schemas and vocabularies that were deemed to meet the following criteria constituted the sample for analysis:

- Freely accessible on the Web;
- Developed to describe information resources;
- Published for use by external institutions; and,
- Published in English (may also be published in other languages).

Different editions of a standard, that were presented as such, were regarded as a single standard with the latest edition being used in the analysis. Integrated sets of individual standards that were presented as such were treated as a single standard.

Assessing candidate standards against the above criteria, particularly (b) and (c), involved a fair amount of judgement. It should be noted that a large number of vocabularies were rejected because although they could be used to describe a particular aspect of information resources, this was not considered their primary purpose. For example, there are many classifications (e.g. of occupations, educational fields, regions, etc.) developed by various statistical bureaux, and these can, and sometimes are, used in biblio-

graphic contexts, but they have been developed, and are maintained, primarily to describe entities other than information resources, such as people. (Information resources were nevertheless defined broadly, to include datasets as well as resources containing "information" or "knowledge.") Thus the compilation differs from those such as BARTOC and ONKI, which do not make this distinction, as well as by those, such as Schema.org and Taxonomy Warehouse, that focus either on schemas or vocabularies. Ultimately, the compilation consisted of 53 schemas and 328 vocabularies; their details are published at http://www.csu.edu.au/faculty/educat/sis/student-resources/lists/information-organisation-element-sets and http://www.csu.edu.au/faculty/educat/sis/student-resources/lists/information-organisation-vocabularies. The size of the compilation is mid-range, in comparison with the metadata registries and directories found elsewhere on the Web; Riley's map featured 105 standards in total.

Each schema and vocabulary, and any accompanying documentation, was examined carefully to ascertain its "coverage" and the (reported) methodology used for its development (if any). The coverage of schemas is defined here with reference to three common (perhaps fundamental) ways of viewing information resources, i.e. their content, carrier and application aspects. Concepts associated with these three aspects were, respectively, subject, form and audience. The last of these may be considered to loosely equate to Riley's "community" variable (2010); reference to "domain" was avoided, due to its ambiguity (Mai 2005). The following preference order was used to classify each schema, in relation to the information resources it is intended to describe (from the evidence examined): their content, then their carriers, then their application. For example, a schema that is for sound recordings in music libraries could be classed in a) music (subject); b) sound recordings (form); or, c) libraries (or more specifically music libraries, i.e. where the resources are used, or at least obtained); according to the order of preference, they would be classed in music. A schema for all kinds and content of sound recordings, on the other hand, would be classed in sound recordings. In the case of the vocabularies, the classifier was left to identify what each of them was intended to describe, which could, potentially, be anything, though typically it could be summarised as a particular "field." General subject vocabularies, such as the *Library of Congress Subject Headings*, were classed according to the ultimate target of their application, i.e. the information resource type (e.g. library materials).

With the emphasis on subject and discipline, the *Dewey Decimal Classification* (*DDC*) was considered a suitable scheme to use to classify the schemas and vocabularies. The scheme is, of course, by no means a perfect or completely universal knowledge organisation system, but it is one of the more general and most widely used in the LIS field, and would be familiar to many of this journal's readers. The classification rules mentioned above took precedence, but otherwise the standard *DDC* rules were applied to number choices, with each standard assigned to only one class. However, no explicit number building was undertaken: the aim was to paint no more than a broad picture of the standards' coverage.

To investigate the (reported) methodologies used to develop both the schemas and the vocabularies, the resource/expert/user warrant typology was used, covering three basic approaches often represented in the LIS literature. The published documentation for each standard was examined to ascertain which of the three warrants, if any, were emphasised in the historical and ongoing development of its semantics. One of the following codes was recorded for each of the standards:

R = resource warrant, i.e. concepts are based on the resources being described
E = expert warrant, i.e. concepts are based on expert guidance
U = user warrant, i.e. concepts are based on users' search needs
C = combined warrant, i.e. methods representing two or more of the above warrants are combined, with no single warrant emphasised
X = methodology is not reported, or unclear, or none of the above applies.

## 4.0 Results

The latest versions of the 53 schemas were published between 1994 and 2014, with a median publication year of 2008. Their nomenclature varied widely; only one was called a "schema," others were "element sets," "specifications," "ontologies," data dictionaries" and so on (the most common designation was simply "metadata standard"). Well-known examples include: *Categories for the Description of Works of Art*, *Dublin Core Metadata Element Set*, *IEEE Standard for Learning Object Metadata* and *SPECTRUM*. The latest versions of the 328 vocabularies were published between 1976 and 2014, with a median publication year of 2010. Several of the standards were called "vocabularies," but many more were called "thesauri" (97 of them, in fact) or "classifications" (27). Well-known examples include: *Art & Architecture Thesaurus*, *ERIC Thesaurus*, *Iconclass*, Library of Congress *Classification* and *NASA Thesaurus*.

### *4.1 Coverage*

A high-level overview of the coverage of the schemas is provided by means of the histogram in Figure 2, showing

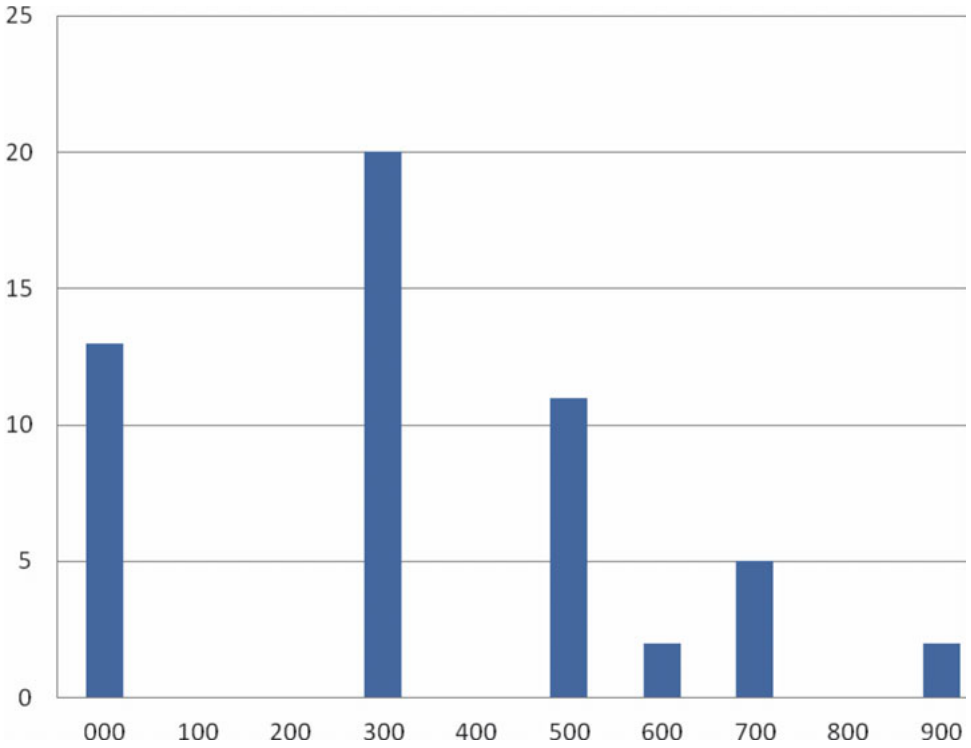


*Figure 2.* Distribution of schemas (main classes)

their distribution across the ten main classes of *DDC* (see Appendix 1 for a list of the ten classes). It should be noted that a lack of coverage cannot be measured quantitatively, as it cannot be assumed that the population of information resources are distributed evenly across *DDC* (indeed, a criticism of the scheme has been that even library resources are distributed unevenly). However, the histogram suggests that more schemas are needed in the areas of philosophy, religion, language and literature, and that technology is also, perhaps more surprisingly, underrepresented. A more detailed account of the schemas" coverage, at the next level down in *DDC*, is shown in Figure 3 (see Appendix 2 for a list of the hundreds divisions). Additional gaps noticeable here are those in law, engineering, manufacturing, fine arts, sport and (non-European) history. Relatively well covered fields include education and publishing.

A high-level overview of the coverage of the vocabularies is shown in Figure 4. The distribution is broadly similar to that of the schemas, with a similar dearth of standards in the areas of philosophy, religion, language and literature. At the next level down, as shown in Figure 5, most subfields in other disciplines are covered to some extent; notable exceptions are construction and some of the fine arts. Relatively well covered fields include many of the social sciences (including law), some of the natural sciences, certain applied sciences, including medicine, engineering and agriculture, and LIS. As one might expect, given their far greater number, the vocabularies cover a larger number of areas than the schemas do, but the areas in which the vocabularies and schemas are most concentrated are broadly similar.

### *4.2 Development methodology*

The extent to which certain types of warrant are emphasised in the development of the schemas and vocabularies in the sample is indicated in Table 1. Documentation covering development methodology was lacking for about half of the standards; documentation was provided for proportinately more of the schemas than the vocabularies. Although a lack of reported methodology does not necessarily mean, of course, that no particular methodology was applied, nor that no internal guidelines were used, an account of how a given standard had been developed would be of use to those considering applying it, as well as for the purposes of this study. The notably high proportion of standards lacking such documentation is not encouraging at a number of levels.

Most of the methodologies reported emphasise a particular warrant. As Table 1 shows, in the case of the schemas, user and expert warrant was reported to be applied much more than was resource warrant; conversely, in the case of the vocabularies, resource warrant was reported to be applied much more than expert or user war rant, with a combination of warrants also applied more than either expert or user warrant alone.

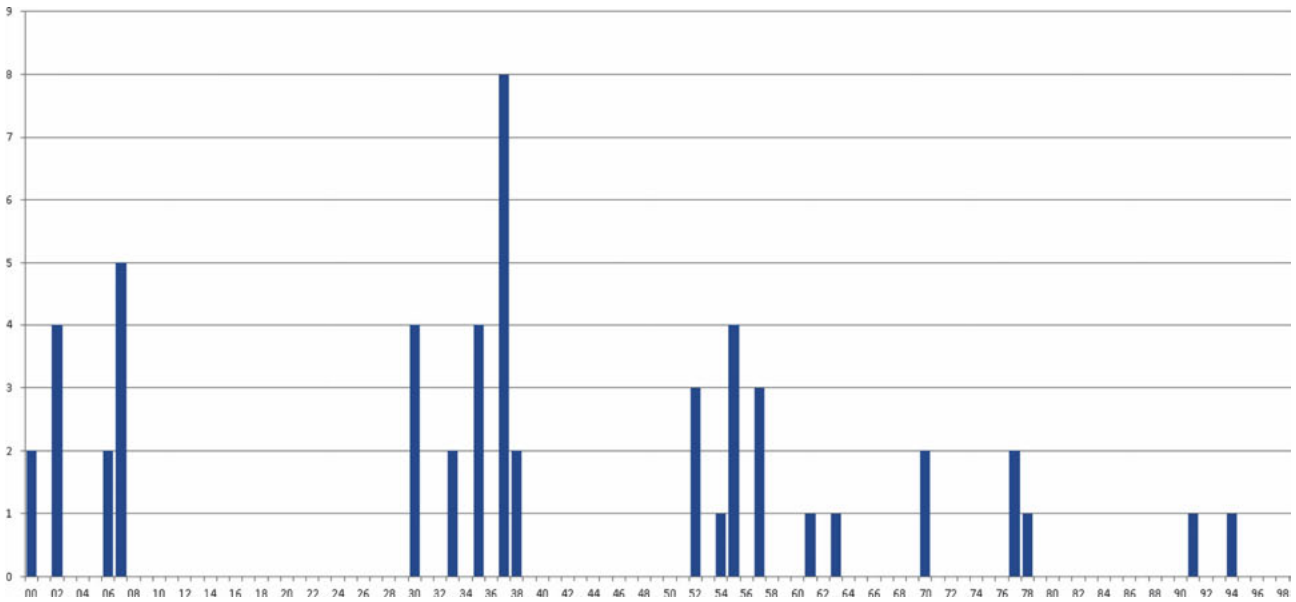

*Figure 3*. Distribution of schemas (hundreds divisions)

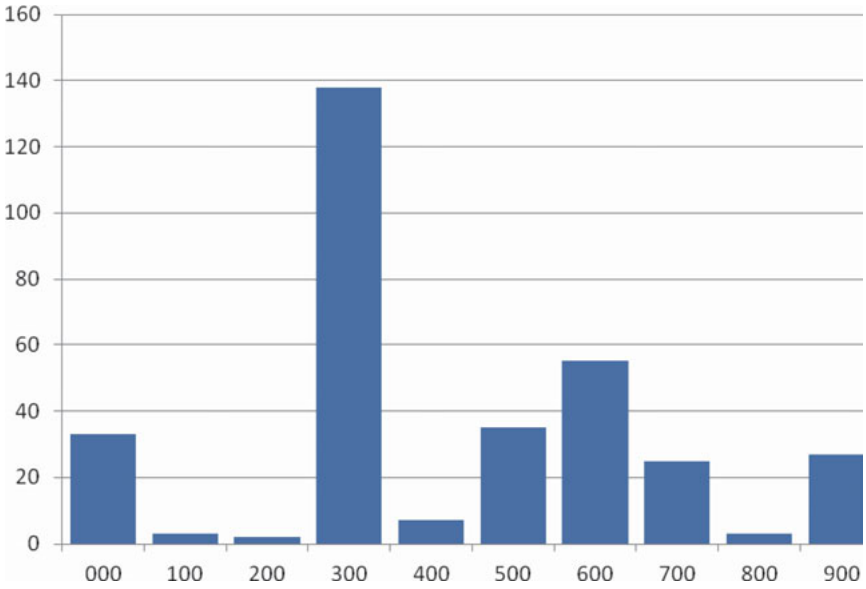


*Figure 4*. Distribution of vocabularies (main classes)

It was speculated that there might also be a relationship between warrant and coverage, so the warrant distributions for the schemas and vocabularies classed in the 300s (social science) and 5-600s (natural and applied science) of *DDC* were compared; they are shown in Tables 2 and 3. For the vocabularies, there are very similar distributions; for the schemas, there is a little more variation, but this could well be due to the small sample size. Thus no relationship between warrant and coverage was discerned.

| **Warrant** | **Schemas** | | **Vocabularies** | |
|---|---|---|---|---|
| | ***n*** | ***%*** | ***n*** | ***%*** |
| **Resource** | 1 | 1.9 | 61 | 18.6 |
| **Expert** | 12 | 22.6 | 18 | 5.5 |
| **User** | 13 | 24.5 | 15 | 4.6 |
| **Combined** | 5 | 9.4 | 42 | 12.8 |
| **Unidentified** | 22 | 41.5 | 192 | 58.5 |
| | 53 | 100.0 | 328 | 100.0 |

*Table 1*. Warrant of schemas and vocabularies

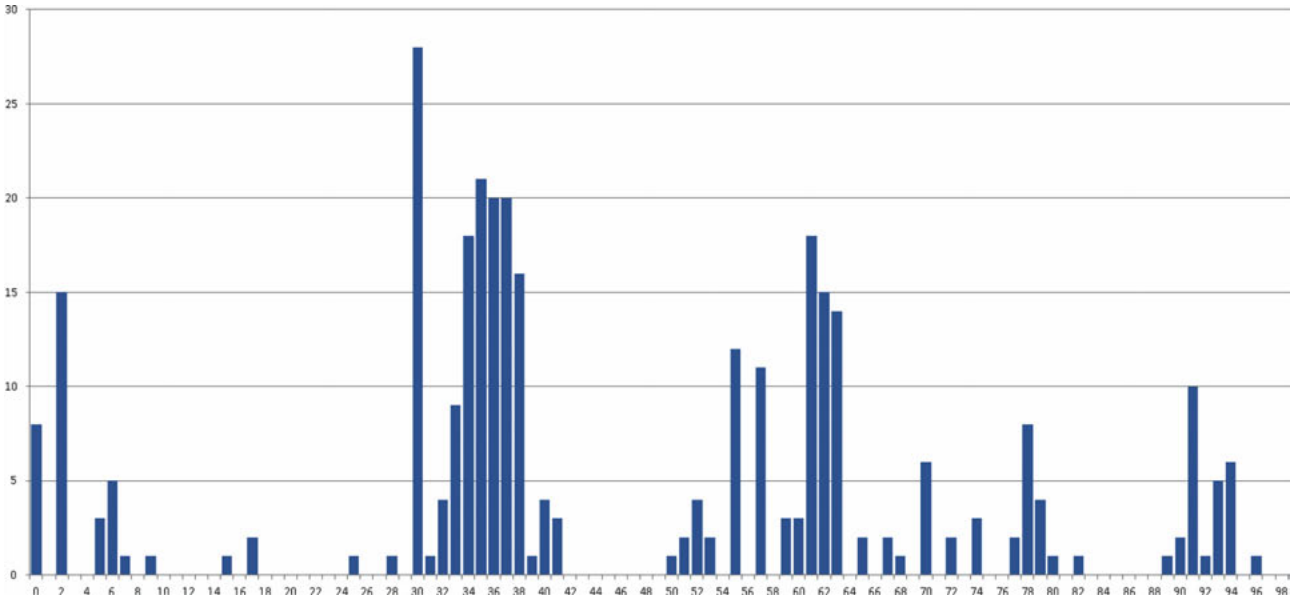

*Figure 5*. Distribution of vocabularies (hundreds divisions)

| Warrant | 300s | | 5-600s | |
|---|---|---|---|---|
| | *n* | *%* | *n* | *%* |
| **Resource** | 0 | 0.0 | 1 | 7.7 |
| **Expert** | 5 | 25.0 | 2 | 15.4 |
| **User** | 3 | 15.0 | 3 | 23.1 |
| **Combined** | 1 | 5.0 | 2 | 15.4 |
| **Unidentified** | 11 | 55.0 | 5 | 38.5 |
| | 20 | 100.0 | 13 | 100.0 |

*Table 2*. Warrant of schemas in *DDC* 300s and 5-600s

| Warrant | 300s | | 5-600s | |
|---|---|---|---|---|
| | *n* | *%* | *n* | *%* |
| **Resource** | 25 | 18.1 | 16 | 17.8 |
| **Expert** | 6 | 4.3 | 6 | 6.7 |
| **User** | 8 | 5.8 | 4 | 4.4 |
| **Combined** | 18 | 13.0 | 10 | 11.1 |
| **Unidentified** | 81 | 58.7 | 54 | 60.0 |
| | 138 | 100.0 | 90 | 100.0 |

*Table 3*. Warrant of vocabularies in *DDC* 300s and 5-600s

## 5.0 Conclusions

The survey results demonstrate that a wide range of schemas and vocabularies are freely accessible on the Web for the description of information resources. No doubt other standards, published but not freely accessible online, extend this range even further, but probably not to an extent that satisfies the needs of all areas. Areas such as the humanities and the fine arts may not need so many schemas and vocabularies, but they probably need more than they currently have; at any rate, they should be targeted for deeper audits. Other fields worth investigating, at least with respect to schemas, include law, engineering, manufacturing and sport. Given that each schema may require the application of multiple vocabularies, more vocabularies than schemas should be needed overall, but how many more, and thus of which type of standard there is a greater shortage, is difficult to judge. The general correlation between the areas covered by the schemas and the areas covered by the vocabularies in the sample suggests that the development of schemas and vocabularies often coincides in practice, as well as in theory.

Assuming the reported methodologies used to develop schemas and vocabularies are broadly representative of those actually applied to develop schemas and vocabularies, there is a notable difference in emphasis: schemas are typically constructed by consulting experts considering end-users' search behaviour; vocabularies, on the other hand, typically focus on the information resources, or combine a range of methods. Why this is the case, and whether it should be the case, remains unclear. It does not appear to be due to differences in coverage, given that similar patterns were reported in both the "hard" and "soft" sciences. It is likely, of course, that the advice offered in vocabulary construction guides such as those cited in section 2.2 above would have had an influence, although they do not fully explain the emphasis on resource warrant; similarly, the case studies that have reported the development of schemas may have influenced other constructions, with an

emphasis on user and expert warrant, but they do not properly explain a relative lack of mixed method approaches to schema development identified in this study.

Also worthy of investigation is the way in which methodologies are combined, when they are. A limitation of his study was that the types of warrant involved in combined approaches were not coded, because relatively low frequencies of such approaches were anticipated and because such granularity would have been harder for the coder to have determined. Optimal development methodology for both schemas and vocabularies is a topic on which much more research is needed. More detailed reporting on the construction of these standards is also to be encouraged.

## References

Aitchison, Jean, Alan Gilchrist and David Bawden. 2000. *Thesaurus Construction and Use: A Practical Manual,* 4th ed. London: Aslib IMI.

Baker, Thomas, Christophe Blanchi, Dan Brickley, Erik Duval, Rachel Heery, Pete Johnston, Leonid Kalinichenko, Heike Neuroth and Shigeo Sugimoto. 2002. *Principles of Metadata Registries: A White Paper of the DELOS Working Group on Registries*. http://web.archive.org/web/20110818100958/http://delosnoe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf.

Beghtol, Clare. 1986. "Semantic Validity: Concepts of Warrant in Bibliographic Classification Systems." *Library Resources & Technical Services* 30: 109-25.

Bliss, Henry Evelyn. 1939. *The Organization of Knowledge in Libraries, and the Subject-Approach to Books,* 2nd ed. New York: Wilson.

Broughton, Vanda. 2006. *Essential Thesaurus Construction*. London: Facet Publishing.

Hider, Philip. 2012. *Information Resource Description*: *Creating and Managing Metadata*. London: Facet Publishing.

Howarth, Lynne, and Eva Jansen. 2014. "Towards a Typology of Warrant for 21st Century Knowledge Organization Systems." In *Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014 Kraków, Poland,* edited by Wieslaw Babik. Advances in Knowledge Organization 12. Würzburg: Ergon Verlag, 216-21.

Hulme, Edward W. 1911. "Principles of Book Classification." *Library Association Record* 13: 354-8, 389-94, 444-9.

Hulme, Edward W. 1912. "Principles of Book Classification." *Library Association Record* 14: 39-46, 174-81, 216-22.

ISO 259641-1. 2011. *Thesauri for Information Retrieval*. Geneva: ISO.

Johnston, Pete. 2004. *JISC IE Metadata Schema Registry: Functions of the IE Metadata Schema Registry*. http://www.ukoln.ac.uk/projects/iemsr/wp2/function.

Kim, Chai and Soon D. Kim. 2007. "Consensus vs Frequency: An Empirical Investigation of the Theories for Identifying Descriptors in Designing Retrieval Thesauri." *Information Processing and Management* 13: 253-8.

Kwaśnik, Barbara H. 2010. "Semantic Warrant: A Pivotal Concept for Our Field." *Knowledge Organization* 37: 106-10.

Lancaster, F. 1986. *Vocabulary Control for Information Retrieval*, 2nd ed. Arlington, VA: Information Resources Press.

Mai, Jens-Erik. 2005. "Analysis in Indexing: Document and Domain Centered Approaches." *Information Processing and Management* 4: 599 -611.

Mai, Jens-Erik. 2008. "Actors, Domain, and Constraints in the Design and Construction of Controlled Vocabularies." *Knowledge Organization* 35: 16-29.

Miller, Steven J. 2011. *Metadata for Digital Collections: A How-To-Do-It Manual*. London: Facet Publishing.

Morville, Peter and Louis Rosenfeld. 2007. *Information Architecture for the World Wide Web,* 3rd ed. Sebastopol, CA: O"Reilly Media.

Riley, Jenn. 2010. *Seeing Standards: A Visualization of the Metadata Universe*. http://www.dlib.indiana.edu/~jenlrile/metadatamap.

Riley, Jenn and Michelle Dalmau. 2007. "The IN Harmony Project: Developing a Flexible Metadata Model for the Description and Discovery of Sheet Music." *The Electronic Library* 25: 132-47.

Rodriguez, Robert. 1984. "Hulme"s Concept of Literary Warrant." *Cataloging & Classification Quarterly* 5: 17–26.

Sayers, W.C. Berwick. 1967. *A Manual of Classification for Librarians*, 4th ed. London: Deutsch.

Souza, Renato, R., Douglas Tudhope and Mauricio B. Almeida. 2012. "Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems." *Knowledge Organization* 39: 179-92.

Zeng, Marcia L. and Jian Qin. 2008. *Metadata*. New York: Neal-Schuman Publishers.

## Appendix 1

000 Computer science, information & general works
100 Philosophy & psychology
200 Religion
300 Social sciences
400 Language
500 Science
600 Technology
700 Arts & recreation
800 Literature
900 History & geography

## Appendix 2

000 Computer science, knowledge & systems
010 Bibliographies
020 Library & information sciences
030 Encyclopedias & books of facts
040 [Unassigned]
050 Magazines, journals & serials
060 Associations, organizations & museums
070 News media, journalism & publishing
080 Quotations
090 Manuscripts & rare books
100 Philosophy
110 Metaphysics
120 Epistemology
130 Parapsychology & occultism
140 Philosophical schools of thought
150 Psychology
160 Logic
170 Ethics
180 Ancient, medieval & eastern philosophy
190 Modern western philosophy
200 Religion
210 Philosophy & theory of religion
220 The Bible
230 Christianity & Christian theology
240 Christian practice & observance
250 Christian pastoral practice & religious orders
260 Christian organization, social work & worship
270 History of Christianity
280 Christian denominations
290 Other religions
300 Social sciences, sociology & anthropology
310 Statistics
320 Political science
330 Economics
340 Law
350 Public administration & military science
360 Social problems & social services
370 Education
380 Commerce, communications & transportation
390 Customs, etiquette & folklore
400 Language
410 Linguistics
420 English & Old English languages
430 German & related languages
440 French & related languages
450 Italian, Romanian & related languages
460 Spanish & Portuguese languages
470 Latin & Italic languages
480 Classical & modern Greek languages
490 Other languages
500 Science
510 Mathematics
520 Astronomy
530 Physics
540 Chemistry
550 Earth sciences & geology
560 Fossils & prehistoric life
570 Life sciences; biology
580 Plants (Botany)
590 Animals (Zoology)
600 Technology
610 Medicine & health
620 Engineering
630 Agriculture
640 Home & family management
650 Management & public relations
660 Chemical engineering
670 Manufacturing
680 Manufacture for specific uses
690 Building & construction
700 Arts
710 Landscaping & area planning
720 Architecture
730 Sculpture, ceramics & metalwork
740 Drawing & decorative arts
750 Painting
760 Graphic arts
770 Photography & computer art
780 Music
790 Sports, games & entertainment
800 Literature, rhetoric & criticism
810 American literature in English
820 English & Old English literatures
830 German & related literatures
840 French & related literatures
850 Italian, Romanian & related literatures
860 Spanish & Portuguese literatures
870 Latin & Italic literatures
880 Classical & modern Greek literatures
890 Other literatures
900 History
910 Geography & travel
920 Biography & genealogy
930 History of ancient world (to ca. 499)
940 History of Europe
950 History of Asia
960 History of Africa
970 History of North America
980 History of South America
990 History of other areas